

CORPUS

**Corpus**

**3 | 2004**

**Usage des corpus en phonologie**

---

## Grands corpus dialectaux ou la phonologie indisciplète

Jean-Philippe Dalbera et Marie-José Dalbera-Stefanaggi

---



### Édition électronique

URL : <http://journals.openedition.org/corpus/269>

ISSN : 1765-3126

### Éditeur

Bases ; corpus et langage - UMR 6039

### Édition imprimée

Date de publication : 1 décembre 2004

ISSN : 1638-9808

### Référence électronique

Jean-Philippe Dalbera et Marie-José Dalbera-Stefanaggi, « Grands corpus dialectaux ou la phonologie indisciplète », *Corpus* [En ligne], 3 | 2004, mis en ligne le 02 décembre 2005, consulté le 19 avril 2019.

URL : <http://journals.openedition.org/corpus/269>

---

© Tous droits réservés

## Grands corpus dialectaux ou la phonologie indiscreète

Jean-Philippe DALBERA\*

Marie-José DALBERA-STEFANAGGI\*\*

\*« Bases, Corpus et Langage », UMR 6039, Univ. de Nice

\*\*« Bases, Corpus et Langage », UMR 6039, Univ. de Corse

**Résumé :** L'article se propose, à partir de l'expérience de la construction et de l'exploitation des bases de données dialectales de la BDLC (corse) et du THESOC (occitan), de cerner ce qu'un grand corpus est susceptible d'apporter à la phonologie. La réponse, appuyée sur quelques cas d'espèces, fait intervenir trois niveaux : celui de l'établissement des faits à soumettre à l'analyse, celui de la validation des hypothèses émises, celui de la valeur heuristique des données prises en compte. Les faits aléatoirement rassemblés dans un grand corpus et donc non nécessairement pris en compte dans l'analyse initiale sont en effet de nature à transformer la vision de la clôture du corpus, à confirmer (ou infirmer ce qui, *in fine*, est aussi important) les hypothèses avancées, à suggérer d'autres solutions aux problèmes rencontrés que celles initialement envisagées. Un des points saillants, en outre, est, au plan théorique, la mise en discussion de la notion d'entité discrète, tant en matière d'unités phonologiques (de type phonème) que d'unités systémiques (de type idiome localisé) tant il est vrai qu'en matière de langue il n'existe qu'un *continuum*. L'idiolecte est déjà composite, l'idiome local n'est qu'un chaînon dans le *continuum* aréal. Les unités discrètes constituent donc l'instrument d'analyse incontournable du fait linguistique, une réalité dans la représentation qu'un locuteur a de son système ; mais l'observable est un *continuum*, un ensemble de cellules perméables les unes aux autres, si c'est bien dans la conscience collective que se trouve (déposée) la langue, conformément au dire saussurien.

### 1. Introduction

Le dialectologue ne pense pas ordinairement à insister sur la nécessité de travailler sur grands corpus tant l'importance

cruciale de ceux-ci lui apparaît comme un truisme. Son quotidien est en effet un corpus multiforme, mouvant, dont la constante est précisément d'être toujours inachevé. Après le recueil et l'analyse des faits du n<sup>ième</sup> parler, il reste toujours le n+1<sup>ième</sup>. De plus, les grands corpus dialectaux se caractérisent par leur prisme large : n'étant pas construits afin de satisfaire un projet particulier, ils ne peuvent, de manière immédiate, se constituer en base empirique de validation d'une hypothèse donnée. Dans le fourmillement de faits qu'ils mettent à disposition, ils sont à même, en revanche, de venir bousculer certaines idées reçues, certaines schématisations excessives, bref de contribuer à une vision moins restrictive de certaines structures linguistiques.

Les quelques pages qui suivent ont pour objet de présenter une esquisse de ce que les grands corpus sont susceptibles d'apporter à l'analyse phonologique, synchronique comme diachronique, qu'il s'agisse d'apport dans le cadre d'une analyse particulière ou d'une contribution plus générale au mode d'approche du niveau phonologique, cette dernière se matérialisant aussi bien sous forme de validation *a posteriori* d'hypothèses que de suggestion pour une progression heuristique. En outre elles sont l'occasion de souligner que les grands corpus dialectaux (le qualificatif est important ici), en relativisant la notion de clôture et en revenant sur le caractère discret des unités ou des systèmes, peuvent contribuer à éclairer et préciser la relation entre apparentements typologiques et parenté génétique et, au delà, à influencer sensiblement sur la modélisation du niveau phonologique de la langue.

Les grands corpus dont il est question dans cet article sont au nombre de deux ; étant du même ordre, ils ont trait à des domaines linguistiques distincts. L'un concerne le corse ; le sigle qui désigne ce corpus est BDLC (*Banque de Données Langue Corse*). Il s'agit au départ d'une base de données destinée à sous-tendre la cartographie du Nouvel Atlas Linguistique de la Corse. Or la possibilité de construire des cartes quasi automatiquement à la demande, c'est-à-dire de livrer des cartes qui sont à la fois non figées et cumulatives, a conféré à chaque objet, la publication papier et l'outil interactif, une certaine autonomie. Dans ce corpus est engrangé

l'ensemble des matériaux recueillis par l'équipe BDLC par voie d'enquêtes de terrain, en Corse au cours des 25 dernières années. L'autre corpus concerne la langue d'oc. Il se nomme *Thesaurus occitan* (THESOC) et il présente sous des dehors comparables un autre type de trésor linguistique : il a en effet pour ambition de rassembler l'ensemble des matériaux de source orale recueillis depuis un siècle environ en matière de langue d'oc et remplissant certaines conditions parmi lesquelles celle d'être localisables. Au premier chef se trouvent donc les *data* réunies dans le cadre du chantier des Atlas linguistiques de la France par régions (CNRS). Or le recueil des faits n'a pas été unitaire dans cette moitié méridionale de la France (les Atlas régionaux ont été pensés en réaction contre le questionnaire unique, l'enquêteur unique, la réponse unique de premier jet afin de corriger les défauts reprochés à Gilliéron) ; il s'est poursuivi sur le temps long (la collection n'est pas achevée à ce jour) ; les auteurs ont imprimé chacun leurs marques et leur style, de sorte qu'il a fallu rendre commensurables, à tous les niveaux, des données qui ne l'étaient pas nécessairement. Ce qui ne peut manquer de sensibiliser à certains problèmes.

Mais les corpus, si grands soient-ils, ne valent que par l'usage qu'on en fait. L'usage que nous évoquons ici est d'ordre dialectologique et peut-être convient-il de préciser ce point. Faire de la dialectologie, c'est, par définition, se confronter non à une langue avec tout ce que cela implique de clôtures, d'usages normés, d'inventaires prétendument finis mais à un ensemble de faits en continuité. La représentation ordinaire des faits dialectaux est soit une collection de monographies soit une mise en perspective diachronique ; dans le premier cas, cela revient à dire que l'idiome localisé est conçu comme l'unité discrète avec laquelle travaille le dialectologue dans sa démarche comparative ; dans le second cas, ce sont des changements qui sont figurés au premier plan, avec leur projection dans l'espace sous la forme matérielle d'isoglosses. De telle sorte que le dialectologue est le plus souvent perçu comme un enquêteur de terrain et un dessinateur d'arabesques isoglossiques dans des espaces où les diachroniciens ont depuis longtemps établi les liens de parenté entre les idiomes. Mais cette représentation du magma dialectal n'est certainement pas

la seule possible ; d'autres globalisations sont envisageables. Il n'est nullement certain, notamment, qu'une comparaison au niveau lexical puisse être conduite sur ces bases. Et, même en restant au niveau phonologique, il est possible que l'effort comparatiste inhérent à la description d'une grappe de parlers conduise à des éclairages non triviaux de certains concepts phonologiques ou des rapports entre typologie et génétique.

La notion de dialecte et, partant, de dialectologie revêtant un caractère ambigu non seulement dans l'usage courant de ces termes mais même dans la terminologie scientifique en vigueur, il convient de préciser, pour prévenir tout malentendu dans la suite du propos, ce que nous entendons par là. Il est d'ordinaire question d'étude dialectale dès lors qu'il s'agit de décrire non une langue officielle mais une variété 'minorée' ou 'locale' ; on ne dit plus guère *patois* (la connotation péjorative ayant été dénoncée par la linguistique) mais on n'en pense pas moins. Or le concept de dialecte ne peut se réduire à cet usage sous peine de perdre tout intérêt ; dans *dialecte* il faut entendre un terme relatif, un terme de parenté : « tel idiome est un dialecte de ... ». L'usage que nous faisons de ce terme est celui qu'illustrent des propositions telles que « le gallo-roman, le gallo-italien, l'ibéro-roman sont des dialectes du roman occidental », ou bien « les langues romanes sont des dialectes du latin ».

Prise dans cette acception, la dialectologie voit son objet se préciser : elle est l'étude de la parenté génétique, plus précisément du degré de parenté génétique entre un certain nombre d'idiomes. Mais cette parenté, là où sa mise en évidence sur la base de correspondances régulières entre les éléments des divers idiomes est déjà acquise, semble pouvoir être établie – structuralisme oblige – en termes de distance génétique entre systèmes. Ainsi, pour donner un exemple trivial, l'auteur d'une reconstruction du système vocalique du roman occidental montre que le système ancien (latin) à deux corrélations (5 voyelles I, U, E, O, A pouvant être longues ou brèves quasiment en toutes positions) se métamorphose sous les effets conjugués de deux tendances générales du phonétisme (que les études typologiques ont établies depuis longtemps) : d'une part la corrélation entre les traits de durée et de tension

des voyelles (/i:/ tend à se réaliser long et tendu [i:] tandis que /i/ tend à se manifester comme bref et relâché [ɪ] etc.), d'autre part l'équilibre quantitatif de la syllabe (voyelle phonétiquement longue en syllabe tonique libre, phonétiquement brève en syllabe entravée). La mise en branle de ces deux principes déclenche une dynamique évolutive qui produit des effets diversifiés selon le degré d'accomplissement de la tendance et les autres changements qu'elle est susceptible de croiser. En d'autres termes, le degré de parenté génétique entre idiomes doit pouvoir s'apprécier en termes de changements s'opérant à partir du système-source et manifestant selon le cas des divergences ou des convergences.

Moyennant quoi, la description linguistique d'un ensemble dialectal se présente comme une étude associant deux perspectives, l'une typologique, l'autre génétique. Typologique, dans la mesure où il convient de donner de chacun des idiomes de l'aire considérée une représentation synchronique, rendant compte de son état au moment du recueil des faits mais intégrant la nécessité de faire apparaître le caractère différentiel de chacun d'entre eux, génétique dans la mesure où le degré de proximité s'apprécie non en différences de structures mais en nombre et nature des changements à partir de la langue-source.

Si l'on peut mener à bien ces deux analyses tout en les gardant soigneusement distinctes et en confrontant leurs résultats, bien des propriétés du comportement des langues, concernant notamment les rapports entre changement diachronique et fonctionnement synchronique appréhendé dans sa variabilité (ici dans l'espace) peuvent être révélées. Les propriétés différentielles qui caractérisent les types mis en évidence se ramènent en effet bien souvent à des systématisations de résultats de certains changements intervenus ou, en d'autres termes, de régularisations de 'désordres' provoqués par ces changements. Les grands corpus dialectaux constituent alors une mine de données inégalable pour permettre de comprendre, à travers les 'issues' multiples que la variabilité spatiale donne à voir, les mécanismes directeurs du fonctionnement linguistique. On a remarqué bien souvent que l'espace, en matière dialectale, apparaît comme une projection du temps : cette formule signifie que bon nombre de

changements qu'un essai de reconstruction conduit à poser, d'ordre phonique notamment et graduellement accomplis sont attestés par tel ou tel idiome ou grappe d'idiomes appartenant à l'aire. Mais la proposition vaut aussi d'un point de vue typologique. Ce que nous appelons l'*innovation partagée* et dont nous faisons le critère fondamental de la partition dialectale et de la définition des aires (Dalbera 1985-86, 1994 et Dalbera & Olivieri 2001), n'est autre, en effet, que l'émergence de types par le biais de généralisations communes consécutives à des situations de déséquilibre structurel. En effet la reconstruction d'un ensemble dialectal consiste certes à établir le système qui sous-tend cet ensemble, mais aussi à comprendre sa dynamique évolutive et à identifier les phases successives de son développement ; cela implique notamment de mettre en évidence la 'pente' phonétique sur laquelle il glisse (ce que, en d'autres temps, on osa appeler des lois phonétiques), de mesurer l'impact de ces changements phonétiques sur le système phonologique et sur l'organisation morphologique, de déterminer la nature et l'ampleur du désordre éventuel induit pour, enfin, mettre en lumière les mécanismes de régulation que chaque système a mis en œuvre pour se rapprocher d'un nouvel équilibre. C'est au cours de cette phase d'innovation que se forment les apparentements dialectaux ; remédier ensemble à un dysfonctionnement de tel ou tel sous-système de la langue, c'est manifester une dynamique de convergence intradialectale et c'est en même temps créer une rupture interdialectale. La part d'héritage (par définition partagée) se distingue alors clairement de la part d'innovation. Or l'évolution phonétique est caractérisée par le continu et le graduel (les choses vont plus ou moins vite, plus ou moins loin, mais sans solution de continuité) ; les restructurations phonologiques ou morphologiques constituent, elles, des ruptures : elles ne concernent pas directement le matériau mais les relations qui le structurent ; ce sont les interprétations des structures linguistiques dans la conscience des locuteurs qui se modifient. Il y a quasi permanence des faits mais modification de certaines des règles qui les régissent ; et, à terme, ces modifications de règles engendrent évidemment des changements dans les faits. Chaque changement d'ordre phonique, quelle que soit son origine, a

pour effet de troubler (au sens neutre, non pas "mettre du désordre" mais "modifier l'ordre") l'économie des idiomes et selon l'entropie induite de susciter une réaction visant à redonner un équilibre aux modules de la langue ayant été perturbés. Cette réaction manifeste la dynamique interne des systèmes et n'est pas univoque : la plupart du temps, la structure de la langue offre plusieurs solutions au problème posé ; c'est là précisément que se mesurent la convergence et la divergence : répondre ensemble c'est se constituer en entité ayant une certaine homogénéité ; remédier différemment aux dysfonctionnements c'est manifester une certaine autonomie et creuser l'écart avec les autres. Il faut ajouter, évidemment, que dans une reconstruction opérant sur le temps long, cette alternance d'entropie et de réorganisation intervient de manière cyclique et n'englobe pas toujours les mêmes idiomes. C'est la raison pour laquelle la vision en à-plat d'isoglosses a pu laisser penser à certains que les frontières (au moins dialectales) n'existaient pas. Mais cela revient, de fait, à ne prendre en compte que les changements et à les mettre tous sur un même plan alors qu'ils sont intervenus successivement et de manière ordonnée. Ce que l'on perd dans cette fusion des calques successifs, c'est toute la part de 'travail' de la langue, de (re)structuration interne qui à la fois définit son identité et esquisse sa dynamique.

C'est dans ce cadre que se situe la leçon que nous avons tirée de la constitution et de l'utilisation des bases de données que nous avons élaborées dans deux domaines dialectaux, la BDLC pour les idiomes corses, le THESOC pour les idiomes occitans. Nous tentons d'en donner ci-dessous quelques illustrations.

## **2. La fonction d'établissement des faits**

Le propos peut paraître paradoxal : on établit toujours les faits ; mais où ceux-ci s'arrêtent-ils ? De toute façon, un corpus n'est jamais assez grand pour comprendre ce tout forcément complexe qu'est une langue. Ainsi une enquête phonologique limitée dans la région du Nord-Ouest de la Corse livrerait des unités comme [paa] "pelle", [a 'a<sup>h</sup>a] "la laine",



[fɛɛ] "foi", [u 'ɛbju] "l'écobuage", [pɛɛ] "pied", [a 'ɛa] "la torche résineuse", [paa] "il paie", [a 'abja] "la cage", [fɛɛ] "fiel", [u 'ɛttu] "le lit", [bɛɛ] "il voit" [a 'ista] "la vue", [a 'ɔttɛ] "le tonneau", [ka'allu] "cheval", etc. Une étude monographique qui se fonderait sur des occurrences de ce type – parfaitement courantes au demeurant, et que l'on peut recueillir, avec une belle constance, auprès de plusieurs locuteurs – s'interrogerait avec perplexité sur le système consonantique de cette variété, la structure syllabique, la fréquence surprenante des hiatus, etc. L'analyse qui en résulterait serait sûrement, en un certain sens du moins, assez peu adéquate à la réalité. La 'réalité' de cette variété, en effet, c'est la coexistence (que l'on ne perçoit qu'en multipliant les informateurs et les occurrences) des formes ci-dessus avec, respectivement, les formes suivantes : [pala], [a 'la<sup>h</sup>a], [f'ɛðɛ], [u 'ðɛbju], [pɛðɛ], [a 'ðɛða], [paga], [a 'wabja], [f'ɛlɛ], [u 'lɛttu], [b'ɛðɛ], [a 'βista], [a 'βɔttɛ], [ka'βallu], 'restituables' par tous les informateurs. Certes le système est en déséquilibre – et les fausses restitutions sont là pour en témoigner – mais, synchroniquement, il fonctionne sur ce double registre, et l'analyse doit en rendre compte.

Les données lacunaires donnent d'ailleurs bien souvent lieu à des interprétations discutables. Ainsi nous recueillons, dans une localité auprès d'un de nos témoins, le terme [viti'dɔɪnu] "sac en peau de veau". Ce terme s'intègre, sans le moindre problème, dans l'organisation phonologique et morphologique du parler tel que nous l'avons analysé jusque là. Il appartient à la classe des masculins à singulier en -u, pluriel en -i. Un autre locuteur (la sœur du précédent) nous donne [viti'dɔɪni] pour le même mot : tout à fait possible aussi d'après l'analyse, mais ainsi suffixé le terme appartient à une autre classe nominale. Les deux informateurs discutent, et pour valider la bonne prononciation font appel à leur mère, qui, elle, prononce [viti'dɔɪ<sup>h</sup>]. Ils sont alors tous les deux pleinement satisfaits : il semble à chacun que la mère a validé sa propre prononciation. Là aussi, on est dans la diachronie en train de se faire : cette région a connu la perte de la voyelle finale après consonne nasale, puis une nasalisation de la voyelle liée à l'affaiblissement de l'articulation de cette consonne nasale. Les voyelles finales après consonnes nasales sont donc des voyelles

restituées et la restitution témoigne d'hésitations. On peut sans doute dire que ni [i] ni [u] n'ont de statut phonologique (ne sont opposables), pour l'heure, dans ce contexte. C'est ce que confirme la situation d'une localité voisine où ces mots présentent régulièrement [ɔ]. La synchronie est décidément un tout bien plus complexe qu'il n'y paraît de prime abord : la variation, même à l'intérieur de ce qui est déjà une variété, ne nous semble pas devoir ni même pouvoir être rejetée, et le corpus doit s'étendre jusqu'à prendre en considération des faits de ce genre, qui sont non seulement des clefs pour l'analyse diachronique mais aussi des pistes éclairantes pour la description synchronique.

Le même type de complexité affleure sur le mode sociolinguistique. Nous avons ainsi découvert, au cours des enquêtes de terrain, des formes éclairantes, par rapport à une hypothèse que nous avons formulée : ['fiɛ] au lieu de ['feɛ] "fiel", ['miɛ] au lieu de ['meɛ] "miel", ['bui] au lieu de ['boi] "bœufs" attendus donnaient des arguments très forts pour appuyer une hypothèse diachronique consistant à interpréter les voyelles mi-fermées comme résultant de la réduction des diphtongues [jɛ] et [wɔ]. Là-dessus, un autre informateur, à ce moment de l'enquête, censure vivement ces formes à voyelles fermées : « on ne parle plus comme ça, ce sont les gens qui parlent mal qui prononcent ainsi », etc. Nous avons eu le même type d'expérience pour les prononciations cacuminales (« on ne doit plus dire ['paɖa] mais ['paʎa] ») ou les dissimilations (['kallu] et non ['karlu]), etc. Ces témoignages sont évidemment précieux.

Mais c'est probablement dans la démarche proprement dialectologique et géolinguistique que cet aspect des choses apparaît dans toute sa vigueur : seul un corpus étendu, au sens spatial, peut donner accès de manière éclairante au fonctionnement d'une langue. La variation, en effet, loin d'être un handicap, constitue une clef incomparable dans la mesure où elle dessine le champ des possibles. Mais pour traquer cette variation rien ne vaut les grands corpus. Si l'on fait référence à l'expérience de la Banque de Données Langue Corse, il faut préciser d'emblée qu'il ne s'agit pas d'un corpus à visée (exclusivement) phonologique. Certes, la BDLC comporte un

sous-corpus, orienté phonologiquement, qui a permis la réalisation du premier volume du *Nouvel Atlas Linguistique de la Corse : Aréologie Phonétique* (Dalbera-Stefanaggi 1995). Ce corpus est fondé sur un questionnaire de 250 unités lexicales qui ont été retenues sur la base de relations synchroniques et d'évolutions diachroniques que les travaux précédents avaient révélé pertinentes. Le résultat des analyses fondées sur ce corpus semble bel et bien valide : on peut définir des clôtures, dans le temps, dans l'espace ainsi que dans le matériau phonique qui donnent accès à ce que l'on appellera, sans trop entrer dans les problèmes, des systèmes phonologiques. Puisqu'il s'agit d'une démarche de géolinguistique, c'est tout naturellement sous forme de cartes (typologiques, diachroniques, de faits isolés ou de systèmes analysés) que les résultats issus de ces corpus trouvent leur meilleure représentation.

Mais la BDLC, c'est aussi tout un ensemble de productions linguistiques plus ou moins contrôlées : enregistrements de textes libres, conversations comprenant de multiples redites, susceptibles de mettre en évidence la variation, chez un même locuteur aussi bien qu'entre locuteurs d'une même variété supposée, parfois transcriptions ou commentaires métalinguistiques des informateurs. Celles-ci ont donné naissance à des hypothèses, le plus souvent relevant de la diachronie, qui se sont parfois avérées fécondes. Ce qui fait un corpus comme celui de la BDLC 'grand', c'est, au fond, la multiplication et le croisement des approches, qui donne accès au polymorphisme. Rappelons la définition que donnait Séguy (1973 : 88) de ce terme : « Le polymorphisme est un aspect du langage tout à fait réel, qu'il faut observer et étudier, qui devra livrer ses secrets, car il est en contradiction avec la loi d'économie. Tout ce qu'on en sait, c'est qu'il fonctionne comme prélude d'un changement à terme plus ou moins long [...]. C'est à la dialectologie [...] que la linguistique doit la découverte de ce champ de recherches captivant ». Dimension constitutive du langage, sans doute. En tout cas, dimension incontournable d'une 'langue en liberté', comme celle sur laquelle nous travaillons, échappant à la norme fixée et explicitée. A ce titre, la BDLC, plus qu'un 'grand' corpus, est un corpus 'épais'. On imagine bien que la constitution du corpus, l'acceptation ou le

rejet de telle ou telle forme posent, de manière aiguë et permanente, la question des critères à appliquer. En domaine dialectalisé et non normativisé<sup>1</sup>, la responsabilité du linguiste est encore plus grande et son exercice encore plus périlleux ; même si l'on convient, comme le dit Bottiglioni, de retenir la « moyenne » de la prononciation des locuteurs, ici et maintenant, on n'est jamais à l'abri de l'erreur d'évaluation ou d'interprétation. Hypercorrection, tentative de s'assimiler au parler d'un autre locuteur ou de l'enquêteur, désir de reproduire la prononciation des anciens, de se conformer à une norme que l'on a plus ou moins entendu évoquer... : les écueils ne manquent pas, et les dictionnaires ne sont pas là pour 'redresser', bien au contraire. De sorte que (et pour notre part, nous y sommes parvenus après quelques années d'incertitude et de réflexion), on gagne peut-être à prendre le parti du flou et du multiple, fort difficile il est vrai à gérer, et à tenter d'explorer les marges, voire les déviances. Ceci bien sûr après avoir établi – ou plutôt tout en tentant d'établir – la « norme », au sens de Bottiglioni (1933-42, préface).

### **3. La fonction de validation des hypothèses**

Les grands corpus présentent également l'intérêt de venir étayer des hypothèses élaborées à partir d'inventaires plus réduits ou plus ciblés : lorsqu'en effet des données 'tout venant' entrent dans les moules qui ont été proposés comme hypothèse sans tenir compte de ces data nouvelles, cela donne à penser que ces moules et, partant, les hypothèses qui en sont à la base, ne sont pas trop inadéquats. Tel ou tel fait anecdotique, imprévisible, qui s'est immiscé dans le tout venant des données d'enquête peut venir confirmer (ou infirmer, mais cela revient au même) une analyse. Ainsi, procédant à l'analyse synchronique du parler de Nice, nous avons été arrêté par l'interprétation à donner de la structure phonologique, apparemment anodine, de mots du type ['kawa] "chose" ou

---

1. C'est délibérément que nous forgeons ce néologisme. Normé, tout système linguistique en usage dans une communauté l'est ; il s'agit alors d'une norme implicite. *Normativisé* renvoie selon nous à une norme prescriptive, un bon usage.

[*'pawa*] "il (se) repose". Notre perplexité était consécutive aux faits suivants : dans ce système a) la diphtongue [*wa*] n'apparaît qu'en position tonique, b) l'élément [*w*] ne dispose d'aucune autonomie en position pré vocalique et ne peut figurer que devant [*a*], c) il existe des diphtongues descendantes à second élément [*w*], parmi lesquelles [*aw*] et [*ɔw*], mais l'opposition qui existe entre ces deux diphtongues en position tonique est neutralisée en position prétonique où seul [*ɔw*] est réalisable<sup>2</sup>, d) [*ɔ*] fait partie de l'inventaire vocalique en position tonique mais ne figure pas en position prétonique. Il s'ensuit que ni la segmentation '*pa-wa*, '*ka-wa* ni la segmentation '*paw-a*, '*kaw-a* ne sont pleinement satisfaisantes, la première parce qu'elle revient à poser soit une unité vocalique /*wa*/ post-tonique tout à fait exceptionnelle, soit une unité consonantique /*w*/ tout aussi exceptionnelle, la seconde parce qu'elle traduit une rupture avec le principe considéré comme 'naturel' selon lequel une chaîne de type CVCV s'analyse prioritairement – c'est-à-dire de manière non marquée – comme CV-CV plutôt que comme CVC-V<sup>3</sup>. On aura remarqué, au passage, que nos deux mots procèdent diachroniquement d'une même configuration source : PAUSA, CAUSA ; et l'on sait que les parlers voisins gèrent cette configuration phonique délicate au moyen d'une épenthèse : [*'pawvɔ*, '*kawvɔ*]. Là-dessus, dans les séquences de discours libre de l'enquête, où il est question des gens de l'arrière-pays – les gavots – nous sommes surpris d'entendre notre témoin les appeler [*gɔ'wat*] et non [*ga'wat*] ou [*ga'vwat*] comme la plupart des locuteurs<sup>4</sup>. La variante, en l'occurrence, 'dit' la règle : le découpage qu'il opère inconsciemment ne peut

2. Par exemple [*'bawba*] / [*bɔw'ba*] "il aboie / aboyer", [*'rawba*] / [*rɔw'ba*] "il vole / voler" mais [*'plɔw*] / [*plɔw'gyt*] "il pleut / plu".

3. Quand il s'agit de la semi-voyelle [*j*], les analystes ne s'embarrassent guère de questions sur le découpage de [*'paja*, '*bɔja*, '*fæja*, '*vjeja*, '*byje*, '*pa'jasa*, '*adʒinu'jat*...] ; celui-ci leur semble être, tout naturellement : '*pa-ja*, '*bɔ-ja*, '*fæ-ja*, '*vje-ja*, '*by-je*, '*pa-jasa*, '*adʒinu-jat* (respectivement "paille", "bourreau", "feuille", "vieille", "il bout", "paillasse", "agenouillé").

4. La source étymologique de ce terme n'est pas complètement claire ; néanmoins il y a tout lieu de penser que le mot appartient à la série que l'on trouve tant en occitan qu'en français : *gavaj*, *gaver*, *jabot*, *goître* ... et procède donc, pour ce qui nous intéresse ici, d'une base \**gab-*.

être que *gaw-at*, seul susceptible de rendre compte à la fois de l'ensemble des faits et des contraintes. Il penche donc vers une interprétation de type CV-V où le noyau vocalique de la première syllabe serait complexe et l'attaque de la seconde syllabe serait vide. Cet exemple saisi au vol – et qui fait apparaître que *ga-vwat* et *paw-zat*, devenus respectivement *ga'wat* et *pa'wat* une fois la consonne intervocalique perdue, sont interprétés de manière analogue – est sans doute une clef majeure pour élucider le phonétisme niçard<sup>5</sup> et il est clair que l'on aurait été bien en peine d'y avoir accès en dehors précisément d'un grand corpus.

De façon un peu comparable, lorsqu'on tente de reconstruire l'évolution du consonantisme des parlers occitans orientaux, on constate que –L latin en position implosive interne connaît, selon les parlers, les locuteurs ou les mots un traitement flottant : tantôt – et c'est la majorité des cas – il aboutit à [–w] (à travers le stade [L]) tantôt il se trouve continué par [R] :

Etymon	contexte /—cs	traitement aire maritime	traitement(s) aire gavotte
ALTA		[ʼawta]	} <i>idem</i>
SALTA (T)	dentale*	[ʼsawta]	
ALTIA (T)		[ʼawsa]	
CAL (Ī) DA		[ʼkawda]	
ALBA	labiale	[ʼawba]	[ʼawba], [ʼalba]
SALVIA		[ʼsawvja]	[ʼsarvja], [ʼsawvja], [ʼsalvja]
ALGA			[ʼarga]
FALCŌNE	vélaire	[fɔw'kuŋ]	[far'kuŋ], [fɔw'kuŋ]
QUALEQUE		[ʼkawke]	[ʼkarke], [ʼkawke], [ʼkalke]

\* que cette dentale soit d'origine ou non.

5. L'un des clivages fondamentaux dans la partition dialectale de l'occitan concerne le sort réservé aux hiatus consécutifs à l'amuïssement de consonnes intervocaliques : aux idiomes conservateurs de structure dans lesquels, par exemple, FĒTA, CĪCER, MATŪRA, RACĒMU ("brebis", "pois chiche", "mûre", "raisin") sont continués par [ʼfea, 'see, ma'yra, ra'in] s'opposent ceux qui 'résolvent' les hiatus : [ʼfjɔ, 'sej, 'ɛjɾɔ, 'rejɱ].

L'hypothèse qui vient à l'esprit est que le changement est contextuel : devant une consonne dentale d'origine ou devenue dentale l'évolution de [l] en [L] puis [w] s'opère ; mais devant labiale ou vélaire cette évolution est bloquée et [l] suit une autre voie. La tendance à généraliser [w] serait alors perçue comme se diffusant à partir de l'ouest (provençal maritime) dans le cadre d'un courant novateur s'inspirant du français 'policé'. De telles hypothèses restent néanmoins fragiles, comme chaque fois que l'on postule des phénomènes de contact et des empilements de strates. Mais les grands corpus livrent les clefs : grands corpus du fait de leur extension aréale (les parlers les plus à l'est attestent clairement le décalage) mais surtout grands corpus du fait des attestations inattendues qu'ils renferment. C'est ainsi qu'une informatrice racontant anecdotiquement un épisode de sa jeunesse à Nice dans le quartier du Port évoque devant nous le personnage d'une parente de notre famille : celle-ci apparaît dans un discours rapporté. Notre informatrice fait état des propos d'une tierce personne s'adressant à elle et, toute à son récit, reproduit la prononciation qui figure de manière marquante dans son souvenir : « madama darbèra ! ... » ; un moment auparavant, cette même informatrice avait raconté qu'elle connaissait plusieurs familles de Dalbera, mais elle avait prononcé [dal'bera]. Il devient évident que ce mot là (patronyme à motif opacifié qui n'a pas de correspondant dans la zone de provençal maritime) ne peut avoir été influencé par un quelconque corrélat prestigieux. C'est bien un indice fort de la partition contextuelle intervenue dans le traitement de -L implosif interne. Mais qui aurait pu imaginer dans son protocole d'enquête et donc de constitution du corpus une question susceptible de faire apparaître ce mot et par suite ce traitement ? Seul le grand corpus, du fait qu'il est ouvert et qu'il fournit des réponses à des questions qui n'ont pas été posées est à même d'étalonner les hypothèses du chercheur.

Dans le même ordre d'idées, les phénomènes d'hypercorrection ou de réfection 'fautive' constituent bien souvent la preuve *a contrario* de la validité des hypothèses diachroniques. Deux exemples : dans les Alpes-Maritimes, la diphtongaison de Ĕ et Ō latins a conduit en position tonique romane dans une série de contextes à l'émergence de

diphthongues ascendantes ; cela, en deux temps : d'abord apparition de [jɛ] et [wɔ] puis, beaucoup plus tard, une fois que [wɔ] a eu évolué en [ɥœ], de [wɔ] de nouveau à partir de [ɔ] dans des contextes phoniques qui n'avaient pas donné lieu à la première diphthongaison. Nous discutons un peu plus loin le statut phonologique de ces diphthongues ; ce qui nous intéresse ici, c'est de voir que le corpus livre des attestations du processus dont nous avons fait l'hypothèse, à savoir que le statut monophonématique de ces diphthongues se délite à partir du moment où des complexes d'autres origines, de type *j* + voyelle, ou *w* + voyelle apparaissent dans la langue et viennent 'concurrencer' les diphthongues ascendantes anciennes. La preuve en est alors apportée, fortuitement, par des évolutions qui attestent le découplage de la diphthongue : dans la masse amorphe du corpus dialectal une forme comme [lu 'ljɛtʃ] "le lit" issue régulièrement de (IL)LU LĚCTU et présentant le même type d'aboutissement que (IL)LU PĚCTU [lu 'pjɛtʃ] "la poitrine" devient [lu 'jɛtʃ] à Roquebillière montrant par là que le groupe initial [l+j] a été interprété /ʎ/ puisque c'est ce [ʎ] (et non la séquence [lj]) qui ultérieurement se réduit systématiquement à [j] comme en témoignent ['fijɔ] "fille", ['pajɔ] "paille", [a'bejɔ] "abeille", [mu'je] "épouse", en face de ['ljewrɔ] "livre", (monnaie), ['ljewkre] "chardonneret"<sup>6</sup>. Mais ce type d'attestation ne peut s'obtenir sur commande ; c'est de nouveau parce que le corpus n'est pas pré-trié et comporte ce que, dans d'autres sphères, on appellerait du tout-venant, que ces attestations-clefs peuvent être répertoriées.

En corse, l'hypothèse que nous avons formulée pour résoudre le délicat problème de reconstruction du vocalisme de type toscan mais caractérisé par rapport au toscan par la

---

6. Bon nombre de parlers attestent encore la réalisation [L] dans la série des continuateurs de FĪLĪA, PALĚA, APĪC(Ů)LA, MULĪĒRE mais non dans [ljewra] < LĪBRA ni dans ['ljasa] forgé sur LĪGARE + suffixe.



fameuse ‘inversion des timbres’<sup>7</sup> s’est trouvée étayée puissamment par un fait anodin qui aurait pu faire défaut à l’analyse si le recueil des données n’avait pas été conçu à grande échelle et de manière non téléologique. Le nom du soc de charrue *passim* [ˈvɔmara], [ˈbɔmara] ou avec l’article [a ˈwɔmara] est attesté en certains points du réseau du grand corpus sous la forme [ˈomara] [l’omara] ce qui constitue une double dérogation aux règles en vigueur : a) le Ō latin (VŌMER) a pour continuateur régulier [ɔ], b) l’initiale du mot étant consonantique, la forme de l’article défini ‘devrait être’, comme c’est le cas dans les autres parlers, vocalique. Tout se passe donc comme si le [wɔ] initial avait été interprété et traité comme un élément vocalique et non comme une séquence v + ɔ et que ce [wɔ] s’était réduit ensuite à une voyelle [o] plus fermée que le [o] qui continue Ū, Ō du latin. C’est précisément ce que prédit l’hypothèse qui postule une diphtongaison des voyelles mi-ouvertes du roman occidental puis une réduction de

7. Par « inversion des timbres » on réfère à une configuration bien identifiée ; Rohlfs (1966 : 9) la décrit comme suit : « In Corsica abbiamo un sistema vocalico a quattro gradi come nei parlari toscani, però con singolare inversione delle antiche qualità ».

(1) Ī	Ū (5)	(6) Ĩ	Ů (10)	(1) i	u (5)
(2) Ē	Ō (4)	(7) Ę	Ö (9)	(2,6) e	o (4, 10)
(3) Ā		(8) Ǻ		(7) ε	ɔ (9)
				(3,8) a	
latin			toscan (= type roman occidental)		
				(1) i	u (5)
				(7) e	o (9)
				(2,6) ε	ɔ (4,10)
				(3,8) a	
			corse centro-septentrional		

Il faut ajouter que la notation phonématique masque une caractéristique importante : les phonèmes /e/ et /o/ se réalisent très fermés [e] et [o] de sorte que les phonèmes /ε/ et /ɔ/ n’apparaissent mi-ouverts que par différence.

ces diphtongues à des éléments mi-fermés, consacrant ainsi l'inversion des timbres<sup>8</sup>.

A partir de quoi la perspective se retourne. Si un grand corpus livre aléatoirement un ou deux témoins d'un mécanisme évolutif important, il devient possible de rechercher des contextes présentant des configurations présumées analogues ; le corpus s'amplifie alors mais cette fois de manière non plus aléatoire mais au contraire concertée. S'ensuivent validation ou falsification de l'hypothèse. Mais dans un cas comme dans l'autre le corpus a joué son rôle.

C'est cette dialectique permanente de la découverte aléatoire<sup>9</sup> et du questionnement construit qui caractérise l'analyse linguistique et qui justifie le grand corpus.

Les exemples qui viennent d'être cités illustrent la validation ou du moins l'étayage d'hypothèses sur des mécanismes qui lient le continuum phonique à la grille d'interprétation phonologique ou à la grille de lecture morphologique. Mais il n'y a pas lieu de limiter là l'apport des grands corpus.

#### 8. Phases présumées :

Ī	Ī Ē	Ē	Ā Ą	Ō	Ō Ū	Ū
	∨		∨		∨	
i	e	ε	a	ɔ	o	u
i	e	jε	a	wɔ	o	u
		jε		wo		
[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]
/i/	/ε/	/e/	/a/	/o/	/ɔ/	/u/

9. Encore faut-il relativiser le caractère aléatoire de la découverte ; cette attestation fortuite ne fait sens que pour celui qui a en tête la question ainsi posée ; aléatoirement signifie ici non pas "par hasard" ou "par chance" mais "de manière non préméditée".

#### 4. La fonction heuristique

En matière d'éclairage dû à l'extension spatiale du corpus, les exemples sont légion. La fragilité des hypothèses étymologiques qui ne s'appuient que sur une forme ou sur un type de formes n'est plus à démontrer. Nous rappellerons simplement ici (bien que le fait ne soit pas à proprement parler d'ordre phonologique) le problème étymologique posé par le terme qui désigne le levain en corse, *nuirme* (cf. Dalbera-Stefanaggi 2001). Salvioni (1916 : 788) se fondant sur l'unique forme à sa disposition voit dans ce mot un continuateur de ALLEVAMEN ; bien des discussions interviennent alors, d'autres étymons sont proposés (NUTRIMEN notamment), tous fondés sur la forme ainsi répertoriée ; mais ces hypothèses deviennent caduques et la solution évidente dès lors que l'on envisage ce terme dans toute l'étendue de sa variation (cf. Dalbera-Stefanaggi 1988) ; une fois l'ensemble des termes recueillis mis en liste (ou en carte), la série : *rruvime*, *rnuivime*, *rinuvime*, *arnuvime*, *nuvirme* ... laisse entrevoir ce qui constitue son identité. Tous ces termes se révèlent être dérivés d'une base commune dont le sémantisme est d'une simplicité remarquable, le "renouvellement" *re-nov-ime* ; toutes les formes à l'abord étrange que revêt cette notion ne sont autres que des formes gravitant autour de *renovime* mais ayant donné prise à toutes sortes de réfections à partir du moment où le motif d'origine s'est opacifié, a cessé de figurer dans la conscience des locuteurs laissant la voie libre à des tentations paronymiques, phonosymboliques ou remotivationnelles. Les détails de cette dynamique ne nous concernent pas ici. Le point pertinent est que la lumière ne surgit que de la confrontation des variantes et que celles-ci n'ont de chance d'apparaître que dans un grand corpus, établi sans déterminisme étroit.

##### 4.1. Corpus et description : la lénition corse

Il arrive que l'amplification d'un corpus et la prise en compte systématique de la variation bouleversent de manière irréversible l'ordre établi par l'analyse. Le fonctionnement du système consonantique du corse peut en donner un exemple.

Dans la situation linguistique corse d'aujourd'hui, le terme *lénition* renvoie avant tout à un mécanisme d'alternance

consonantique initiale, phénomène repérable parce qu'il concerne la variation du signifiant. En fait – nous l'avons montré ailleurs (Dalbera-Stefanaggi 2001) – il s'agit là de la mise en relation synchronique de deux processus qui ne sont que très partiellement liés : le mécanisme de *renforcement consonantique de l'initiale* et le mécanisme de *lénition* qui s'opère – éventuellement – sur cette même initiale, quand elle se trouve en position intervocalique au sens large, c'est-à-dire dans un certain nombre de contextes non 'renforçants'<sup>10</sup>. Ainsi, les locuteurs établissent une relation stricte entre *trè ppani* vs *u bane* et *skappadu* vs *skabadu* par exemple (respectivement "trois pains", "le pain" et "échappé", "étêté"). Le mécanisme assume une importante valeur sociolinguistique, cette sonorisation et l'alternance initiale qui y est corrélée apparaissant comme une originalité à l'égard de l'italien standard, qui est la référence à partir de laquelle doit explicitement s'appliquer la règle de l'écart maximal. Les enquêtes de terrain révèlent le rôle de marqueur de ce phénomène, en particulier chez les locuteurs jeunes, du fait de leur apprentissage largement scolaire, d'une part, et de leur attitude de militantisme linguistique, d'autre part, cela potentialisé par l'émergence d'une *koinè* (cf. Dalbera-Stefanaggi 2001). En effet, si les jeunes locuteurs appliquent, de manière quasi systématique, la règle de sonorisation des consonnes sourdes intervocaliques, les relevés effectués auprès de locuteurs plus âgés font apparaître une situation infiniment plus nuancée. Il convient par conséquent d'aller au-delà d'une vision schématique de la situation de la Corse au regard de la lénition consonantique – vision du reste non dénuée de fondement – qui consiste à opposer une aire septentrionale largement sujette à la réalisation sonore des consonnes sourdes intervocaliques, à une aire méridionale nettement plus conservatrice, présentant un début de lénition portant essentiellement sur les constrictives et sur le couple k/g. La réalité est en effet plus complexe, tant dans les modalités de la lénition que dans sa source et dans sa répartition aréologique.

---

10. Sont, en revanche, renforçantes les finales consonantiques (ou à consonne sous-jacente) et/ou oxytoniques, comme [tr'ɛ] < TRÈS.

Les modalités tout d'abord : nombre de ces lénitions ne sont pas forcément des sonorisations mais peuvent prendre la forme d'une spirantisation, associée ou non à une sonorisation. Ainsi, sur *capelli* "cheveux", [i ɣa'bɛlli], [i ɣa'βɛlli], [i ja'βɛlli] à côté de [i ga'bɛlli], "les cheveux". La source – ou plutôt les sources – ensuite : il faut distinguer les traces de lénition romane au sens strict, en l'occurrence gallo-italique, dont l'origine septentrionale, en Toscane comme ailleurs, ne semble pas faire problème : il s'agit d'une série bien identifiée de termes septentrionaux intégrés au dialecte toscan et véhiculés ensuite jusqu'en Italie méridionale : *lago* "lac", *pagare* "payer", *spada* "épée", *strada* "route", *povero* "pauvre", *vescovo* "évêque", *padella* "poêle", *scodella* "écuelle", *gridare* "crier" ... Ces lénitions sont dites « lexicalisées », dans la mesure où ce qui doit être supposé comme formes de base, ce sont ces formes *lago*, *pagare*, *strada* (et non les formes latines : LACU, PACARE, etc.) ; cela seul permet de rendre compte des formes modernes : *strada* > ['strada], tout comme *coda* > ['kɔda] – ou encore ['stra<sup>ɔ</sup>a] et ['kɔ<sup>ɔ</sup>a], voire ['straa] et ['kɔa] aujourd'hui, dans l'ensemble de la Corse (exceptée l'aire du cap-corse, on y reviendra), en fonction du degré qu'a connu l'application de la nouvelle lénition. Car la Corse est par ailleurs partie prenante dans le cadre d'un autre phénomène de lénition – celui sur lequel prend appui la démarche moderne d'élaboration linguistique : dans le domaine italo-roman, ce phénomène connaît une diffusion centro-méridionale, plus précisément latio-ombrienne. Les manifestations de cette lénition dans le domaine italien rappellent largement ce que l'on relève en domaine corse. En Toscane, l'imbrication entre cette lénition et la spirantisation est grande, et ne va pas sans poser, sur le plan génétique, un certain nombre de problèmes. Il semble indispensable de distinguer entre lénition comme mécanisme – c'est le cas présenté par l'Italie centro-méridionale – et lénition comme résultat : la lénition de la Romania Occidentale ne concerne que la position intervocalique au sens strict, alors que, en Italie centrale et méridionale, le phénomène se produit également à l'initiale en position intervocalique (phénomènes de phonétique syntactique), entraînant une variation du

signifiant, ce qui rend précisément le phénomène perceptible en synchronie<sup>11</sup>.

Le tableau ci-dessous présente les phonèmes consonantiques des systèmes corses centro-méridionaux<sup>12</sup> :

p̄, t̄, k̄, c̄, k̄ <sup>w</sup>	['skāpu] "je m'échappe", ['fātu] "fait", ['tāku] "talon", ['māca] "maquis", ['āk̄ <sup>w</sup> a] "eau"
p, t, k	['kapu] "tête", ['fata] "fée", ['aku] "aiguille"
b, d, ʝ, g	['babu] "père", ['pedi] "pied", ['āʝu] "j'ai", [p'agu] "je paie"
f, s, ʃ	['pāfi] "fesses", [p'āsi] "pas", ['fāʃi] "langes"
v, z, ʒ	['muvra] "mouflon", ['muzu] "museau", ['bāʒu] "baiser"
t̄s, t̄ʃ	['pātsu] "fou", ['fāt̄ʃu] "je fais"
t̄ʃ	['fat̄ʃi] "il fait"
d̄ʒ, d̄z	['pad̄ʒina] "page", ['med̄zu] "milieu"
m̄, n̄, r̄	['māma] "mère", ['pāni] "linge", ['tāra] "terre"
m, n, ɲ, l, r	['amu] "hameçon", ['mani] "main", ['māɲa] "il mange", ['palu] "pieu", ['karu] "cher"
β (ou w), j	['fāβa] "fève", [ma'jɔ] "grand"

auxquels s'adjoint, dans l'extrême-sud, d̄ : ['pād̄a] "paille".

Les parlers de l'aire septentrionale se caractérisent, sur le plan du consonantisme, du moins en première approximation, par la réduction du nombre de degrés d'opposition entre p̄, p et b, k̄, k et g, t̄ʃ, t̄ʃ et d̄ʒ etc :

['skāpu], ['kabu], ['babu]  
['tāku], ['agu], ['pagu]  
['fāt̄ʃu], ['fad̄ʒi], ['pad̄ʒina]

Il s'agit de l'effet, en synchronie, du phénomène diachronique de lénition consonantique. Les incidences sont

11. On est là en présence d'une variation dont la 'responsabilité' incombe à la syntaxe, dans la mesure où celle-ci opère de manière différente sur un niveau phonologique uniforme, selon qu'elle permet ou ne permet pas l'application du mécanisme à travers la frontière de mot.
12. Inventaire en position intervocalique. Dans ce qui suit nous notons, par convention et sur la base d'analyses conduites par ailleurs (Dalbera-Stefanaggi 1991) les consonnes 'fortes' surmontées d'un trait horizontal, qui indique leur longueur.

d'ailleurs souvent encore plus lourdes puisque, selon les consonnes et les aires considérées, l'amuïssement peut aller jusqu'à son terme ; c'est le cas pour la dentale sonore, qui peut varier de [d] à zéro, en passant par diverses variantes affaiblies : [pɛdɛ], [pɛδɛ], [pɛ<sup>δ</sup>ɛ], [pɛ<sup>w</sup>ɛ], [pɛ<sup>l</sup>ɛ], [pɛɛ].

Pour en revenir aux parlers du centre-sud, qui offrent l'état le plus 'complet' du système, une vision plus large permet d'intégrer, à condition de prendre en considération la position initiale à la frontière de morphème, les phonèmes suivants (positions dites initiale faible, initiale forte) ; ainsi : c : [rica'ra] "rincer", k<sup>w</sup> : [a k<sup>w</sup>ali'ta] "la qualité", ġ<sup>w</sup> : [a ġ<sup>w</sup>ān'ta] "attraper", ĵ : [pa ĵoku] "par jeu", ts : [a 'tsaḡa] "la bêche", etc.

L'opposition fort / faible qui structure l'ensemble du consonantisme corse sous-tend le mécanisme d'alternance consonantique initiale. L'initiale du mot est en effet marquée, en corse, par une alternance consonantique qui se réalise, schématiquement, selon les règles suivantes : la consonne initiale se trouve soit en position forte, soit en position faible. La position faible se définit par "précédée d'une voyelle atone" ; la position forte est constituée de tous les autres cas : précédée d'une voyelle tonique, d'une pause, d'une consonne ... Cette complémentarité de distribution entraîne, pour une unité donnée, deux variantes, selon le contexte après lequel cette unité figure ; ainsi le mot *surella* "sœur" : [a zu'rella] "la sœur", [tɾɛ ssu'relle] "trois sœurs" ; ou *cane* "chien" : [d'ui 'gḡni] "deux chiens", [u<sup>n</sup> 'kkḡne] "un chien". Cette répartition épouse bien entendu l'organisation phonologique consonantique de chaque aire. A titre d'exemple, voici sa réalisation dans l'aire septentrionale :

- pp/b : [tɾɛ 'ppḡni] "trois pains" / [u 'bḡne] "le pain"
- tt/d : [in 'ttæra] "par terre" / [a 'dæra] "la terre"
- kk/g : [a 'kkɔlpi] "à coups" / [d'ui 'gɔlpi] "deux coups"
- cc/ĵ : [ɛ 'ccara] "c'est clair" / [unn'ɛ 'mmikka 'jara] "ce n'est pas clair"
- bb/w : [a 'bbiṣtu] "il a vu" / [anu 'wiṣtu] "ils ont vu"
- dd/δ : [un 'ddɛnte] "une dent" / [d'ui 'δɛnti] "deux dents"
- gg/w : [i<sup>n</sup> 'ggɔla] "dans la gorge" / [di 'wɔla] "de gorge"
- ĵ/j : [pɛr 'jogu] "par jeu" / [u 'jogu] "le jeu"

*Grands corpus dialectaux ou la phonologie indiscreète*

- g<sup>w</sup>/w : [i<sup>h</sup> 'g<sup>w</sup>æra] "en guerre"/ [a 'wæra] "la guerre"  
gr/r : [ɛ 'grānu] "c'est du blé"/ [u 'rānu] "le blé"  
mm/m : [a 'mmamma] "à ma mère"/ [di 'mamma] "de ma mère"  
nn/n : [ɛ 'nnadu] "il est né"/ ['era 'nadu] "il était né"  
ll/l : [a 'lluni] "à lundi"/ [u 'luni] "le lundi"  
tts/dz : ['sɔ 'ttsekki] "ce sont des tiques"/ ['una 'dzekka] "une tique"  
ttj/ddʒ : ['sɔ ttj'imidʒɛ] "ce sont des punaises"/ ['una 'dʒimidʒa] "une punaise"  
ff/v : ['sɔ 'ffɔɛ] "ce sont des contes"/ ['parenu 'vɔɛ] "on dirait des contes".  
ss/z : [ɛ 'ssɔlu] "il est seul"/ ['era 'zɔlu] "il était seul".

Ce mécanisme d'alternance consonantique, et l'opposition des degrés de force consonantique, sur lequel il s'articule, présents dans l'ensemble de la Corse, ne se réalisent cependant pas partout avec autant de netteté. L'affinement du maillage d'enquête et l'amplification du corpus apportent régulièrement des restrictions à l'étendue du mécanisme. On a ainsi pu mettre en évidence (Dalbera-Stefanaggi 1995 carte 245) une aire composée de deux ailes distinctes et reliées que nous avons appelée *aire de réaction à la lénition consonantique*. La carte ainsi conçue globalise des synthèses partielles qui se fondent sur des faits qui ont semblé convergents.

Tout d'abord se dégage, en zone-tampon entre l'aire septentrionale et l'aire méridionale une aire qui se caractérise par l'assourdissement de certaines occlusives sonores intervocaliques, là où la zone méridionale présente des sonores : ['bapu] "père", ['sapatu] "samedi", ['acu] "j'ai", ['pocu] "éminence" (Dalbera-Stefanaggi 1995 cartes 47, 57, 161, 178, 238, 244...). La distribution, pour des raisons diachroniques évidentes, est lacunaire, et sans doute il ne s'agit que de vestiges. Mais la relation de cette aire avec la partie reliée et distincte qui lui fait écho semble significative. Dans cette partie (il s'agit de l'aire du Taravu), nous avons eu beaucoup de difficultés à établir les faits. Les premières enquêtes, sur la base du questionnaire phonologique général, donnaient des résultats qui semblaient s'intégrer globalement dans l'aire méridionale,



mais des contre-exemples de distribution étaient frappants. Nous avons multiplié les points d'enquête, ils ont confirmé... l'incertitude. Un informateur, qui possédait de toute évidence une excellente intuition phonologique, nous a alors confié des textes, qu'il avait rédigés dans une graphie spontanée, et l'incertitude d'audition s'est muée en certitude d'interprétation : ces parlers possèdent, en variantes 'libres', une occlusive bilabiale (mais les choses se vérifient, *mutatis mutandis*, pour les autres consonnes) qui peut varier de *pp* à *b*, en passant par *p* et *bb*. Les variations de graphie concernant une même unité tout comme les incohérences diachroniques relevées ne pouvaient s'expliquer qu'ainsi. On relève par exemple *lubbu*, *luppu* "loup" < LUPU, *sapatu*, *sappatu* "samedi" < SABBATU, *cabbu*, *cappu* "tête" < CAPUT etc...ou encore *pegi* "poix" < PICE, *viccinu* "voisin" < VICINU, etc. Nous avons alors repris systématiquement tous les enregistrements libres réalisés dans cette localité et nous nous sommes rendu compte que le même phénomène (variation très étendue du degré de force de la consonne) se réalisait également à l'initiale du mot, au mépris de la règle (établie par la doxa) de l'alternance consonantique initiale : on peut trouver aussi bien *u ppani* que *trè pani*.

#### 4.2. Les révélations du corpus

Cet écrasement des degrés de force consonantique – délicat à gérer en typologie synchronique – est peu à peu apparu comme un véritable révélateur des lignes d'évolution, comme un principe explicatif fort. Dans un premier temps, il a éclairé un des problèmes posés par la 'rupture' vocalique qui se situe au sud de cette ligne : dans l'aire de l'extrême-sud, qui connaît un système consonantique fondé sur les degrés de force, on trouve un système vocalique qui, si l'on s'en tient à des bases diachroniques, possède cinq phonèmes en position tonique (issues de la confusion, terme à terme, des longues et des brèves latines). Les voyelles moyennes connaissent une variante fermée en syllabe ouverte et ouverte en position entravée (le système possède des géminées, que l'analyse nous a conduits à ne pas poser pour les autres aires (cf. Dalbera-Stefanaggi 1991). On a ainsi ['pedi] "pied", mais ['pettu] "poitrine", ['ʃɛltu] "choix", ['soli] "soleil" mais ['nɔtti] "nuit", ['mɔrtu] "mort".

L'organisation consonantique décrite pour la zone du Taravu n'a que faire de la complémentarité entre [e] et [ɛ], [o] et [ɔ] devant une consonne variable. On aura donc, en fonction de la nouvelle donne, [ˈpedi] "pied", [ˈpettu] "poitrine" mais [ˈʃɛltu] "choix" ; [sˈoli] "soleil", [ˈnotti] "nuit", mais [ˈmɔrtu] "mort".

Dans un second temps, cette écoute orientée a été étendue à d'autres parties du corpus déjà classées : il est possible de mettre en évidence tout un ensemble de vestiges, y compris dans d'autres corrélations consonantiques, qui témoignent du fait que, diachroniquement, la distribution fort / faible qui prévaut actuellement à l'initiale, n'a pas toujours été la règle : çà et là, et en particulier en limite des aires qui appliquent le plus massivement la lénition, on trouve *unu nnumaru* "un nombre", *due mmane* "deux mains", *unu nnum'iku* "un ennemi"... Plus généralement, des traces de renforcement de consonnes présumées faibles sont réparties dans toute la bande centrale de la Corse (cf. Dalbera-Stefanaggi 1995, carte 243).

Les scories de l'analyse phonologique synchronique deviennent dès lors des clefs diachroniques. La répartition aréologique, revue à cet éclairage, loin de correspondre à la bipartition schématique déjà évoquée, épouse des contours plus complexes qui témoignent d'une histoire tourmentée. La question peut être abordée par la situation du cap-corse, marge de l'aire marginale que représente la Corse. En premier lieu, celui-ci se caractérise par une quasi absence des phénomènes de lénition de type centro-méridional, c'est-à-dire des lénitions qui affectent, selon les mêmes modalités, les consonnes intervocaliques et les consonnes initiales en position faible (cf. Médori 1999, Dalbera-Stefanaggi 2001, III.7). Excepté chez les jeunes générations, marquées par la diffusion du modèle évoqué plus haut, peu d'occurrences de consonnes sonorisées ou spirantisées dans ces contextes. En second lieu, il se révèle ne pas connaître la lénition romane « lexicalisée » commune à la Corse et à la Toscane, et diffusée jusqu'en Italie méridionale : on y relève en effet [ˈstrata] "route"... où le degré de l'occlusive latine est conservé. Eventuellement, ces formes subissent la « nouvelle lénition » : [ˈstraɖa] ou [ˈstrada], qui supposent aussi la base latine et non la base au consonantisme

déjà affaibli – qui aboutirait à [ˈstraða] ou [ˈstraa] – que l’on doit poser pour le reste de l’aire. En troisième lieu un phénomène original de renforcement consonantique, inconnu ailleurs en Corse, a été récemment mis en évidence dans l’aire cap-corsine (cf. Médori 1999) : il concerne, schématiquement, l’initiale du morphème après préfixe *s-* (ou *dis-*), et s’applique tant aux occlusives sonores (qui s’assourdissent) qu’aux constrictives (qui deviennent occlusives). On a ainsi : [ˈbɛrka] "barque", mais [spɛrˈka] "débarquer", [ˈfilu] "fil", mais [spiˈlatu] "effiloché"<sup>13</sup>. Or ce phénomène, on le sait (cf. Dalbera-Stefanaggi 1991, §§218-223) est indépendant, dans sa genèse, du phénomène de lénition, avec lequel il est ultérieurement connecté pour donner lieu au mécanisme moderne déjà décrit. L’aire cap-corsine se présente donc ainsi comme témoin d’un état qui présente non pas la lénition, mais le renforcement consonantique initial, base d’organisation du mécanisme d’alternance déjà décrit. Cette organisation, dans le domaine dialectal italien, renvoie aux dialectes de l’Italie méridionale (cf. sur ce point Rohlfs 1966-69).

La mise en relation de certains faits de l’ordre de ceux qui précèdent a conduit on l’a vu, à établir ce que nous avons appelé une *aire de réaction à la lénition consonantique* (cf. Dalbera-Stefanaggi 1995, carte 245). Cette aire est composée de sous-aires, au demeurant assez disparates par ailleurs, mais qui ont en commun la propriété suivante : non seulement elles ne présentent pas de consonnes affaiblies, par rapport à la source, mais encore elles présentent des ‘fortes’ là où on attendrait des ‘faibles’. Si la vision première qui s’impose, en matière de lénition comme sur bien d’autres plan, est celle d’une bipartition qui suivrait plus ou moins la ligne de partage des eaux (*cismonte* vs. *pumonte*), une vision plus attentive – et, surtout, des relevés de micro-dialectologie, qui permettent d’exhumer des strates plus anciennes sous un habillage

---

13. On ne saurait éliminer complètement l’hypothèse d’une simple assimilation de voisement associée à un ajustement destiné à ‘éviter’ un groupe *sf-* mais d’autres faits plaident en faveur d’un mécanisme de renforcement, en contexte post-consonantique (cf. notamment Médori 1999).

moderne – tracent une image un peu différente. L'aire de « nouvelle lénition », celle que l'on connaît en Italie centro-méridionale, ainsi que dans les aires latérales de la Toscane, opère une large trouée, dont l'épicentre est situé sur la côte orientale, un peu au sud de Bastia, et se trouve entourée d'aires sans lénition, de « réaction à la lénition ».

Ce que nous entendons souligner est tout autre : c'est la nécessité de prendre en compte, dans la description et l'analyse, la variation de la langue, variation qui, malgré le flou qu'elle semble engendrer, est au contraire fort significative et bien structurée. Seul un corpus 'épais', évolutif, protéiforme, semble de nature à permettre de rendre compte de cette dimension fondamentale du langage.

## **5. L'impact sur la modélisation : langue et unités discrètes**

### **5.1. Des phonèmes non-discrets?**

La variation apparaît donc comme un révélateur majeur des structures linguistiques ou de l'évolution de celles-ci. Mais la question se pose de savoir quel statut donner à ce polymorphisme omniprésent dans le temps, l'espace et la société. Les choses ne sont pas simples.

En phonologie, tout semble affaire de discrétion : phonèmes discrets, systèmes clos, donnant forme, dans la démarche qui nous intéresse, à des partitions géolinguistiques bien ourlées. Les transitions s'exercent dans des aires (géographiques) et selon des contextes (phoniques) et sociaux (âge des locuteurs) aisément appréhendables. Les zones intersticielles elles-mêmes se laissent appréhender à travers ce prisme : entre l'aire à 7 phonèmes vocaliques en position tonique et celle à 8 phonèmes dans la même position se trouve tout un espace d'hésitation structurée, où les variantes se distribuent en fonction des contextes, de l'âge des locuteurs, de divers paramètres fluctuants avant de se fixer en unités phonématiques discrètes.

De telle sorte qu'on est porté à dire que si les anciens n'ont pas trouvé la loi, au sens phonologique du terme, c'est qu'ils n'ont pas raisonné en termes de systèmes mais de réalisations concrètes (cf. Bottiglioni 1933-42, *Introduzione* :

139-149). Bottiglioni, en effet, n'assortit pas sa démarche géolinguistique et diachronique d'analyses ponctuelles<sup>14</sup> : quels que soient les rapports qui existent entre les continueurs de telle ou telle unité, ces continueurs n'en sont pas moins, en synchronie, des unités qui font système avec d'autres. Mais en réalité, dans un corpus 'en temps réel', perpétuellement en train de se constituer, les choses sont loin d'être toujours claires, malgré un filtre de lecture contraignant. Dans les cas de ce type, c'est la profusion de termes et de variantes de ces termes fournie par le corpus qui permet d'appréhender la phonologisation comme un processus en marche, ou complètement abouti.

Mais parfois, dans ce type de situation, le polymorphisme est tel qu'il rend l'analyse délicate et périlleuse : il est la source de deux incertitudes majeures dans la description, l'une relative à la détermination même des unités phonématiques, l'autre relative au choix des traits distinctifs de ces dernières (Dalbera-Stefanaggi 1991 : 220-222). Le premier problème est celui des chevauchements de réalisations. Il se pose – très schématiquement – de la manière suivante : deux sons [x] et [y] apparaissent, ou bien chez le même locuteur ou bien selon les locuteurs, à l'intérieur d'une zone géolinguistique homogène, tantôt en opposition, tantôt en variation, sans que le contexte phonique y soit pour quelque chose. C'est le cas par exemple lorsque, à l'intérieur d'une série d'unités lexicales, se manifeste une unité qui connaît une certaine marge de variation (de [x] à [y]) tandis que par ailleurs, à l'intérieur d'une autre série d'unités lexicales, [y] se manifeste encore, mais de manière stable, cette fois, sans être sujet à variation. Il s'ensuit que dans certains cas la relation entre [x] et [y] est une opposition (on peut opposer des séquences de type [axb] à [ayb]), tandis que dans d'autres une unité lexicale donnée peut se réaliser indifféremment [cxd] ou [cyd], [x] et [y] apparaissant comme des variantes libres. L'organisation du système est alors pour le moins fuyante. Donnons un exemple, emprunté au parler de Sampolu, zone du Taravu (sud de la Corse). Les

---

14. On est évidemment, dans les années 20, dans une ère pré-phonologique et on ne saurait reprocher à Bottiglioni de raisonner en termes pré-structuralistes !

locuteurs de cette localité distinguent des unités comme ['fɛdi] "foi", ['pɛri] "poires" et ['pɛli] "poils". Mais le champ de dispersion de chacune des unités qui se trouvent représentées à l'intervocalique est étendu à un point tel qu'il peut recouvrir, en partie, celui de l'unité voisine. Ce que l'on peut convenir de noter /d/ est réalisable non seulement comme [d] mais aussi comme [ɔ̃] ou encore comme un [r] à un seul battement. Ainsi, à côté de ['fɛdi] "foi", on trouve des réalisations ['fɛɔ̃i] ou ['fɛri]. Ce qu'on peut convenir de noter /r/ connaît une variation qui peut le conduire de [ɔ̃] à [r] ou à un [ɹ] légèrement vibré : à côté de ['pɛri] "poires", on rencontre ['pɛɔ̃i] et ['pɛɹi]. Ce que l'on peut convenir de noter /l/ enfin est susceptible de varier de [r] jusqu'à [l] : à côté de ['pɛli] "poils", s'entendent ['pɛri] et ['pɛɹi]. Le recouvrement n'est jamais total et les réalisations ne sont pas interchangeable : par exemple /d/ peut se réaliser éventuellement [r] mais /r/ ne peut pas se réaliser [d]. De sorte que l'on n'est pas en présence d'une 'simple' variation mais plutôt d'une 'chaîne d'oppositions', /d/ vs. /r/ vs. /l/, au moins potentielles et toujours actualisables. On pourrait représenter de la manière suivante la coexistence entre ces oppositions, et le brouillage qu'elles subissent du fait de leurs réalisations partiellement communes :

/d/			/r/			/l/		
[d]	[ɔ̃]	[r]	[ɔ̃]	[r]	[ɹ]	[r]	[ɹ]	[l]
1	2	3	2	3	4	3	4	5

La seconde incertitude introduite par ce genre de situation est relative, nous l'avons dit, au choix des traits susceptibles d'opposer ces phonèmes et, par là même, à leur classification. Lorsque, entre deux unités corrélées, sont susceptibles de jouer, simultanément ou alternativement, plusieurs distinctions phoniques, laquelle retenir comme trait distinctif? Et dans le cas de champs de dispersion étendus avec chevauchements, tout choix de trait distinctif n'est-il pas abusif? Est-il légitime, par exemple, de définir comme occlusif le [d] de Sampolu, sachant qu'il peut aussi bien se réaliser [ɔ̃] ou [r]? Et, dans le cas contraire, quel est le trait qui serait susceptible d'assurer la distinction entre /d/ et /r/, /d/ et /l/? On se trouve constamment au seuil de la contradiction, et il faut insister, dans la mise en évidence de ce polymorphisme, et de

ses interprétations multiples, sur le fait qu'on se trouve en présence de plusieurs systèmes superposés, chacun se définissant probablement d'ailleurs par une certaine hiérarchisation des traits distinctifs. Elaborer une typologie des systèmes consonantiques consistera par exemple non seulement à présenter divers systèmes et à mettre en lumière la répartition de ceux-ci dans l'espace, mais aussi à appréhender, à travers chaque système, les diverses couches qui s'y superposent. C'est à ce prix seulement que, nous semble-t-il, on pourra se faire une idée de ce tout complexe que constitue l'état synchronique.

### ***5.2. Des degrés de phonématisation ?***

Les faits occitans tels que les révèle un grand corpus ne sont pas en reste dans cet examen critique des analyses ponctuelles confrontées à la diversité. En matière d'unités discrètes notamment et de modélisation de la structure phonologique. Combien de fois le linguiste – et cela, nous osons dire, indépendamment du cadre théorique qu'il s'est choisi – s'est-il trouvé devant des faits embarrassants ne 'collant' pas aux représentations proposées par les modèles en vigueur ! Combien de fois la 'solution' à laquelle il s'est rallié lui a-t-elle semblée fragile voire réversible et, partant, donnant l'impression de forcer la réalité ! Mais 'forcer la réalité' est peut-être une formulation impropre, car la réalité dont il est question ici est essentiellement l'interprétation implicite que les locuteurs se donnent de leur propre langue : les faits, eux, restent les faits.

Si l'on considère, par exemple, le cas du statut des réalisations [jɛ], [ɥɛ], [wɔ] déjà évoqués plus haut dans les idiomes constitutifs de l'aire dialectale des Alpes-Maritimes, on se trouve en face de la situation suivante : ces diphtongues ascendantes figurent dans tous les idiomes ; mais elles sont tantôt strictement confinées à la position tonique, tantôt étendues à d'autres positions dans le mot ; elles sont tantôt clairement en opposition l'une avec l'autre (susceptibles d'apparaître dans les mêmes contextes phoniques) tantôt distribuées de manière complémentaire ; elles sont tantôt partie prenante d'alternances systématiques, tantôt sujettes très épisodiquement à ces alternances ; les semi-voyelles qui précèdent le noyau vocalique

sont tantôt indissolublement liées à la voyelle qui les suit, tantôt susceptibles de se combiner quasi librement à n'importe quelle voyelle. Imaginons que quelqu'un s'attelle à la description de l'un des idiomes présentant ces diphtongues dans une distribution qui ne serait systématique ni restrictivement ni positivement. Quel que soit le plateau de la balance qu'il déciderait de faire pencher, et aucune modélisation connue, à notre sens, ne l'en dispenserait, il garderait l'impression que sa description va au delà des faits. Et ce malaise ne se dissipe, de fait, qu'au moment où la prise en compte d'un grand corpus fait apparaître de manière évidente que l'espace donne à voir, à travers le continuum dialectal, des degrés de phonématisation. Mieux même, la conduite puis la confrontation de deux analyses réalisées l'une dans une perspective de typologie synchronique, l'autre dans le cadre d'une reconstruction permet de se rendre compte que 'la meilleure hypothèse' au plan diachronique, qui tend à concevoir les choses comme une série de diphtongaisons conditionnées des voyelles brèves mi-ouvertes d'origine ne constitue nullement 'la meilleure hypothèse' pour rendre compte du fonctionnement synchronique à époque moderne des systèmes d'alternances et que tout se passe comme si ces diphtongues, désormais autonomes, venaient à se réduire, par perte de leur élément le plus ouvert, à des voyelles simples lorsqu'elles figuraient en position prétonique.

## **6. Conclusion : la langue en tant que *continuum* de systèmes communicants ?**

Ce qui est décisif, nous semble-t-il, dans l'utilisation de grands corpus pour mener à bien des analyses phonologiques c'est, paradoxalement, la possibilité d'exploiter la variation. On conçoit généralement celle-ci comme une complication inutile dont on se passerait volontiers, susceptible d'alourdir les règles, d'engendrer des exceptions aux mécanismes que l'on a supposés. Il en est de fait ainsi à une certaine échelle, lorsque notamment on a décidé la clôture nette d'un objet. Mais la variation peut aussi être envisagée comme un atout de première grandeur lorsqu'on se propose de saisir la structure d'un



ensemble d'idiomes. Rendre compte au plan phonologique d'un groupe de parlers en synchronie, cela implique de faire une série d'analyses mais ne se limite pas à cela ; il faut encore dégager dans cette liste inerte de systèmes une partie stable (nous voulons simplement dire par là partagée par l'ensemble des idiomes) et faire apparaître les caractéristiques de chacun d'entre eux. Ce que l'on vient d'indiquer comme base commune pourrait représenter un diasystème ; ce diasystème serait une entité synchronique permettant d'étalonner chacun des idiomes particuliers. Dans cette perspective, les idiomes apparaissent comme autant de variantes de ce diasystème, autant de manifestations concrètes d'une forme abstraite. Le diasystème serait, vu de cette manière, à la démarche typologique ce que le protosystème est à la démarche de reconstruction. Les lignes de force de la typisation, qui façonnent les contours des idiomes sont autant de garde-fous pour les hypothèses de reconstruction.

Quelle que soit la manière de regarder une langue, celle-ci présente un caractère de *continuum*. La discontinuité, en matière de langage, est vraisemblablement un prisme interprétatif aussi bien de l'usager de la langue (et dans ce cas le prisme est implicite) que de l'analyste de la langue (et dans ce cas le prisme est explicite et prend la forme d'une théorie linguistique).

Nous avons confronté tout d'abord corpus clos / corpus ouvert pour aboutir à la conclusion que la clôture n'était pas dans les faits mais dans le prisme d'appréhension de ceux-ci. La chose est patente et peu originale. Nous avons ensuite établi que les faits décisifs pour orienter un système d'hypothèses n'étaient peut-être pas toujours accessibles de manière immédiate mais pouvaient se nicher dans des espaces marginaux. Nous avons enfin non pas mis en doute mais relativisé la notion d'unités discrètes, en phonologie en montrant que le polymorphisme dans certains cas était de nature à rendre vaines les définitions de certaines unités phonématiques. Mais qu'est-ce au juste que ce polymorphisme ? C'est simplement la traduction concrète du fait que les locuteurs ne sont pas monolingues ou plutôt ne possèdent pas un système unique ; ils disposent notamment de

plusieurs registres : prononcer ['paa] mais être capable à tout moment de revenir à la forme 'complète' ['pala], cela témoigne du fait que le locuteur gère deux normes, l'une disons de cercle restreint (ou intime ou familial, peu importe l'étiquette) et une autre plus large, et qu'il est capable en fonction du contexte d'énonciation de se servir de l'une ou de l'autre ou d'alterner les deux : il n'y a donc pas solution de continuité entre les usages. La description du système idiolectal et du système microcommunautaire ne sont pas pour autant invalidées mais elles sont tributaires d'un prisme préalable. Il en va de même, *mutatis mutandis*, des parlers : dans un espace géographique, où passe la limite entre le parler *a* et le parler *b* ? à partir de quel moment sait-on que l'on ne parle plus la langue *a* mais que l'on est dans la sphère de *b* ? Il en va de même encore dans un espace temps : à partir de quel moment quitte-t-on une synchronie pour entrer dans une autre ? Formulée ainsi la question prête à sourire : la synchronie est un concept opératoire non une réalité matérielle. Quelle que soit la dimension dans laquelle on se situe, il n'y a que des *continuum*.

Synchronie, monographie, idiolecte, idiome sont des notions assurément nécessaires au plan opératoire mais n'ont guère d'existence et restent des artefacts au plan référentiel.

Mais s'il est vrai que même la phonologie peut-être indiscreète, du fait que tout système reste susceptible de communiquer avec un système adjacent, il faut peut-être concevoir le module phonologique des langues de manière comparable au module lexical, en ce sens que, comme lui, il est commandé par un mécanisme cyclique : les règles qui gèrent les séquences phoniques et qui façonnent de manière spécifique le matériau phonique de chaque idiome ont une durée ou une aire d'application limitée : pendant une période donnée ou à l'intérieur d'un espace donné, elles exercent une prégnance puis, pour de raisons diverses, s'amortissent et finissent par perdre tout pouvoir actif ; mais cela n'implique nullement que les matériaux qu'elles ont façonnés soient soumis à réfection. Ceux-ci continuent à fonctionner dès lors du fait de la convention d'usage en tant qu'éléments figés : ils sont désormais lexicalisés. Sort comparable à celui des unités significatives qui participent d'un cycle de trois phases :

motivation à la création, inhibition (possible) du motif une fois le relais pris par la convention d'usage, arbitrarité enfin pour peu que le motif ait été complètement opacifié ; et de même qu'il y a des degrés dans l'opacité / transparence des unités significatives, il y a des degrés dans le filtrage des séquences phoniques. Les grands corpus nous invitent à dépasser certaines clôtures et à penser le fonctionnement linguistique en termes de macrosystème.

### **Grands corpus cités**

**BDLC** : Banque de Données Langue Corse , base de données multimédia, sous la direction de M.J. Dalbera-Stefanaggi, Université de Corse-CNRS UMR 6039.

**THESOC** : Thesaurus occitan, base de données multimédia, programme interuniversitaire (Nice, Aix-Marseille, Toulouse, Lyon) sous la direction de J. Ph. Dalbera, Université de Nice-CNRS UMR 6039 (ce programme a bénéficié des soutiens du CNRS à travers l'ILF, de la DGLF, des Régions PACA et Midi-Pyrénées, de la Communauté Européenne, de l'AIEO et du Ministère de la Recherche à travers l'ACI SHS-SPI.

### **Références bibliographiques**

- Bottiglioni G. (1933-42). *Atlante Linguistico Etnografico Italiano della Corsica*, Pisa.
- Dalbera J.-Ph. (1985-86). « Alpes-Maritimes dialectales. Essai d'aréologie », *TCLN* 7-8 : 3-28.
- Dalbera J.-Ph. (1994). *Les parlers des Alpes-Maritimes. Etude comparative, essai de reconstruction*. Londres : AIEO.
- Dalbera J.-Ph. (1996). « Strates et représentations dans une base de données dialectales », in G. Moracchini (éd.) *Actes du Colloque International « Bases de données linguistiques : conceptions, réalisations, exploitations »*. Corte, pp. 103-116.

- Dalbera J.-Ph. (1996). « La base de données dialectales du Thesaurus Occitan », *Bollettino dell'Atlante Linguistico Italiano*, III<sup>ème</sup> série. Torino, pp. 187-202.
- Dalbera J.-Ph. (1998) « La base de données Thesoc. Etat des travaux », in J. Gourc & F. Pic (eds.), *Toulouse à la croisée des cultures*, II. Toulouse, pp. 403-417.
- Dalbera J.-Ph. & Oliiviéri M. (2001). « Reconstruction et modélisation. Les parlers de l'aire mentonnaise », *TCLN* 20 : 109-141.
- Dalbera-Stefanaggi M.-J. (1995-99). *Nouvel Atlas Linguistique de la Corse*, I –II. Paris : CNRS Editions.
- Dalbera-Stefanaggi M.-J. (1988). « Evolution phonétique et démotivation : le 'levain' corse », *Espaces romans I*. Grenoble : Ellug, pp. 205-212.
- Dalbera-Stefanaggi M.-J. (1991). *Unité et diversité des parlers corses. Le plan phonologique. Parenté génétique et affinité*. Alessandria : Edizioni dell'Orso.
- Dalbera-Stefanaggi M.-J. (2001). *Essais de linguistique corse*. Ajaccio : A. Piazzola Editions.
- Medori E. (1999). Les parlers du Cap Corse. Etude de microdialectologie. Thèse de doctorat Univ. de Corse.
- Rohlfs G. (1966-69). *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino : Einaudi.
- Salvioni C. (1916). « Note di dialettologia còrsa », *Rendiconti, Reale Istituto Lombardo* I—XLIX, fasc. XVI. Milano : Hoepli, pp. 705-880.
- Séguy J. (1973). « Les atlas linguistiques de la France par régions », *Langue Française* 18 : 65-90.